Dynamics of RNA Translocation through a Nanopore

Julius B. Lucks[1] and Yariv Kafri[2]

[1]*Lyman Laboratory of Physics, Harvard University, Cambridge, MA 02138*
[2] *Department of Physics, Technion, Haifa 32000, Israel.*
(Dated: February 9, 2008)

We present a simplified model of the dynamics of translocation of RNA through a nanopore which only allows the passage of unbound nucleotides. In particular, we consider the disorder averaged translocation dynamics of random, two-component, single-stranded nucleotides, by reducing the dynamics to the motion of a random walker on a one-dimensional free energy landscape of translocation. These translocation landscapes are calculated from the folds of the RNA sequences and the voltage bias applied across the nanopore. We compute these landscapes for 1500 randomly drawn two-letter sequences of length 4000. Simulations of the dynamics on these landscapes display anomalous characteristics, similar to random forcing energy landscapes, where the translocation process proceeds slower than linearly in time for sufficiently small voltage biases across the nanopore, but moves linearly in time at large voltage biases. We argue that our simplified model provides an upper bound to the more realistic translocation dynamics, and thus we expect that all RNA translocation models will exhibit anomalous regimes.

I. INTRODUCTION

The translocation of polynucleotides through nanopores has been recently studied extensively, both experimentally and theoretically, due to its relevance to important biological processes [1, 2, 3] (see below) and emerging sequencing technologies [4, 5]. The standard experimental setup consists of a nanopore embedded in a thin membrane that separates two buffered solutions (see Figure 1 for a schematic.) Initially the cis-side contains the polynucleotides of interest, and when a voltage is applied across the membrane, the electric field couples to the negatively-charged backbone of the polynucleotide to provide a driving force for translocation into the trans-side

The great interest in this system stems from the small dimensions of the nanopore. The most commonly used nanopore is the α-hemolysin (αHL) pore derived from *Staphylococcus aureus*, which at its smallest diameter is 1.5 nm [6, 7]. The presence of such a narrow pore hinders the passage of double-stranded polynucleotides which arise naturally when, for example, RNA molecules fold on themselves due to attraction between the bases. Synthetic nanopores, in which the pore is carved out of a synthetic membrane such as silicon nitride, are studied as well [5, 8] and offer the advantage of a range of pore sizes that can be synthesized for application-specific purposes.

In cases where only single-stranded polymers can pass, nanopore translocation makes an ideal model system suited to studying fundamental biological processes in a controlled fashion. Indeed, biological processes that involve translocation of a polynucleotide across a narrow pore include bacteriophage infection [9], RNA helicases [10], and most importantly the ribosomal protein synthesis machinery [11, 12, 13]. As ribosomes move along messenger RNA (mRNA), any loops that have formed due to folding of the mRNA must be broken in order for protein synthesis to continue [1]. It is well known that messenger RNA secondary structure can cause a frameshift in the decoding of the mRNA into protein [2], and secondary structures that hinder mRNA translocation through the ribosome are considered to be another facet of the complex gene regulation machinery employed by cells. Experimentally, this has been investigated on only the simplest systems of single hairpins [14], while theoretical studies have focused on detailed investigations of specific sequences [15, 16, 17]. We comment that even without any secondary structure of the mRNA, the dynamics of ribosomes can be rich due to the sequence heterogeneity of the mRNA track and the different chemical fuels used by the ribosome [18].

In contrast to the rather poor understanding of the translocation of structured polymers, much is known about the translocation of unstructured polynucleotides through nanopores. Several experiments have focused on the translocation of single-stranded DNA (ssDNA) or ssRNA homopolymers which cannot bind on themselves [4, 5, 6, 7, 8, 19, 20, 21, 22, 23, 24]. When the polymer enters the nanopore, it restricts the flow of the buffer ions in the solution and causes a current blockade. By measuring the duration of the blockade, the experiments can measure the distribution of translocation times. In general, these distributions have been shown to be bimodal as a result of the assymetry in translocating either the 3', or 5' end first [23, 25, 26].

The theoretical work on the translocation of unstructured polymers through nanopores is significant and has focused on the configurational entropic barriers to translocation as a result of the pore confinement [27, 28, 29, 30], the role of the interaction of individual base pairs with the pore [23, 25], and other details such as the viscous drag caused by moving the negatively charged polymer and its counterions through solution [25]. There has also been significant effort in various simulations including simplified models [26, 29, 30, 31, 32] and full-atom simulations [23].

Most interestingly, there has been discussion concerning regimes of polymer length and applied voltage where the

dynamics of translocation become anomalous [30, 33]. By anomalous dynamics we mean that either the velocity, v, or the diffusion constant, D, or both are not well-defined in a limiting sense [34]. Operationally we can define velocity with the limit $v = \lim_{t \to \infty} \frac{x(t)-x(0)}{t}$, which is trivial when the position, $x(t)$, is governed by a linear relation with time, t, $x(t) = vt$. However, in anomalous dynamics, the position can grow as t^μ with $\mu < 1$, a sublinear powerlaw of time, or even slower which makes $v \to 0$ in this limiting sense. Alternatively, when the position grows as t^μ, the typical time it takes a polymer of length N to cross the pore scales as $N^{1/\mu}$. It is important to identify likely candidates for anomalous dynamics, not only to appreciate why the dynamics may be so slow, but also to help the interpretation of experimental data [33] since an experimentally measured velocity becomes dependent on the time-window used in the measurement [18].

In this work, we focus on understanding the dynamics of translocation of *structured* polynucleotides such as ssRNA or ssDNA that can bind on themselves in a complementary fashion. In particular, we focus on how the requirement of base pair unbinding at the nanopore affects the dynamics of translocation. To this end, we ignore the configurational entropic barriers, nucleotide-pore interactions, and other details that were important in considering unstructured nucleotides. These simplifications are motivated by the fact that, as we show, the fluctuations in the base pairing energies grow as $N^{1/2}$, where N is the length of the translocating ssRNA, while other barriers typically grow logarithmically [30]. Furthermore, experimental observation show that unstructured single-stranded nucleotides translocate approximately 100 times faster than hairpins which must be first unzipped to translocate [35]. Thus any base-pair opening requirements are the rate limiting step for the translocation dynamics and we are safe to focus our attention on them.

We show that while the folded structure of the RNA can be rather complicated, the dynamics of translocation behave very similarly to those of a random walker on a random forcing energy landscape (RFEL). Namely, there exists a regime of voltages where the number of translocated bases, m, grows sub-linearly as a function of time with $\overline{\langle m(t) \rangle} \sim t^\mu$ with $\mu < 1$, where $\langle \ldots \rangle$ denotes a thermal average while the overline denotes an average over realizations of the RNA sequence.

This paper is organized as follows: In Section II, we outline the theoretical study of RNA translocation through nanopores by deriving a Free Energy Landscape (FEL) in analogy with DNA unzipping. Section II A discusses important differences between DNA unzipping and RNA translocation in terms of the branched nature of the RNA translocation process. We also discuss a simplifying assumption used in this work to enable us to examine only one-dimensional FELs. Section II B describes how these one-dimensional landscapes are calculated for RNA translocation. We present results of the landscapes and model dynamics performed on the landscapes in Section III, and conclude with a discussion of how these dynamics can be explained in terms of random forcing energy landscapes, and implications for experiments in Section IV.

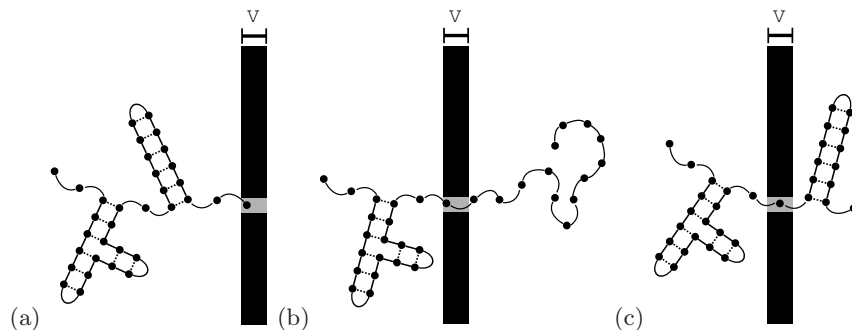## II. MODEL AND ENERGY LANDSCAPES OF RNA TRANSLOCATION

FIG. 1: Model translocation experiment. (a) A negatively-charged single stranded polymer (RNA or ssDNA) starts out on the cis (left) side of a nanopore, which divides a chamber filled with buffer solution. The polymer is a linear string of randomly chosen letters from either a 2 (A,U) or 4 (A,U,C,G) letter alphabet that allows complimentary pairing (dotted lines) (A pairs to U and C pairs to G) and can thus fold on itself (U is replaced by T for DNA). A voltage bias, V, applied across the nanopore (grey) embedded in a membrane (black) provides a driving force for the polymer to translocate to the trans (right) side of the pore where it can be forced to remain unbound to itself (b) (no-refolding), or rebind on itself (c) (refolding). The diameter of the pore only allows single-stranded polymers to pass, and thus all base pairs on the cis-side must be broken in order to translocated.

Figure 1 depicts a schematic of a typical translocation experiment. A membrane divides a buffer-filled chamber into cis and trans sides, and initially a single stranded polymer (representing either RNA or equivalently ssDNA) is placed in the cis-side. The RNA molecule has a negatively charged backbone and is composed of a one-dimensional string of randomly chosen nucleotides consisting of the four bases (A,U,C,G) (with U replaced by T for ssDNA). The nucleotides can pair with each other in a complimentary fashion in which A pairs to U (T) and C pairs to G, making the polymer on the cis-side able to fold on itself as depicted in Figure 1(a) [15]. In what follows we refer to a particular matching of the bases as a fold of the RNA molecule.

Embedded in the membrane is a nanopore which is the only pathway between the cis and trans sides. A voltage applied across the nanopore provides the driving force necessary for the negatively-charged polymer to translocate to the trans-side. We assume, similar to the αHL nanopore, commonly used in experiments, that the diameter of the pore only allows single-stranded polymers to pass: a base pair that is blocking translocation at the pore in the cis-side must be broken before translocation can proceed, an event which requires overcoming the free energy of the base pair bond (see Figure 1(a)). If this occurs then the fold of the RNA molecule on the cis-side must readjust to accommodate the loss of a base to the trans-side.

For the bases that emerge on the trans-side there are two possibilities : (a) they can either be forced to remain unbound or (b) if the specific configuration allows they can rebind creating a fold of the RNA molecule on the trans-side (Figure 1 (b) and (c) respectively.) The former will be referred to from here on as the no-refolding case, and the latter as the refolding case. Both of these cases are examined in detail below. In studying the no-refolding case, we primarily intend to disentangle the complications of refolding on the trans-side from the translocation barriers imposed by the unfolding on the cis-side, rather than mimic an actual experimental setup. However, there are several ways to enforce the no-refolding condition including the addition of single-stranded binding proteins on the trans-side [36], or attaching a bead to the trans end of the RNA, which is pulled by laser tweezers [15]. Both methods would add additional complications to our model below, and are not discussed further.

To describe the dynamics of translocation we consider the free-energy change as the RNA molecule is threaded through the pore base-by-base. As the RNA translocates, the secondary structures changes in order to achieve equilibrium. The dominant configuration of the RNA in equilibrium, at experimental temperatures, is close to the lowest energy secondary structure, given the constraints imposed by the pore. In reality, if the change in configuration required to reach the ground state is radical enough, the molecule does not reach this state [37]. This makes the problem inherently non-equilibrium in nature, and either requires sophisticated numerics, or simplified models [16]. In this work, we focus on the latter, with the objective of obtaining a simple picture on which more complicated, realistic models can build.

With this in mind, our model assumes that the system has enough time to reach its (possibly degenerate) ground state at each step of the translocation process. We expect our results to hold even if this assumption is not strictly valid, as long as the RNA configurations are close to the ground state. A similar approach has been applied successfully to the study of translocation of *unstructured* single-stranded polymers through nanopores, where the assumption is validated both through arguments concerning rate-limiting timescales in the problem, and through comparison to experimental results [25, 26, 38].

Under these assumptions the free energy change as a single base pair is translocated has several contributions, with the most important for *structured* polynucleotides being (a) the free energy gained due to the voltage bias, (b) the free energy required to unbind the base on the cis-side (if it was bound) and (c) the free energy gained due to any rebinding of the base on the trans-side (if possible). There are additional free energy barriers to translocation including entropic barriers due to physical polymer configurations [27, 28]. However, under the equilibrium assumption mentioned above, both of these entropic barriers contribute logarithmic terms to the free energy [29, 30], which can be neglected compared to the much larger barriers of breaking and reforming base pairs [15] for long enough RNA molecules. These barriers, as we show, give contributions to the free energy that grow as $N^{1/2}$ where N is the length of the translocating molecule.

Even with the simplification of just including base pair breaking and reforming, along with the free energy gained due to the voltage bias, the free energy landscape that describes how the free energy changes as each base is translocated is a highly non-trivial structure. We examine this next, and construct a simplified *one-dimensional* landscape which we use to study the dynamics of the translocation process.

A. Branched, Non-Equilibrium Translocation Landscapes

Consider first a simple RNA molecule of length N with a hairpin fold in the no-refolding scenario. (Figure 2(a)). The first $N/2$ bases are A and are complementary to the last $N/2$ U bases, causing all bases to be paired in the initial fold, when the molecule is completely on the cis-side. A sufficient voltage applied across the nanopore provides the driving force necessary to unzip the hairpin to enable translocation [14, 35]. Since the cooperativity parameter
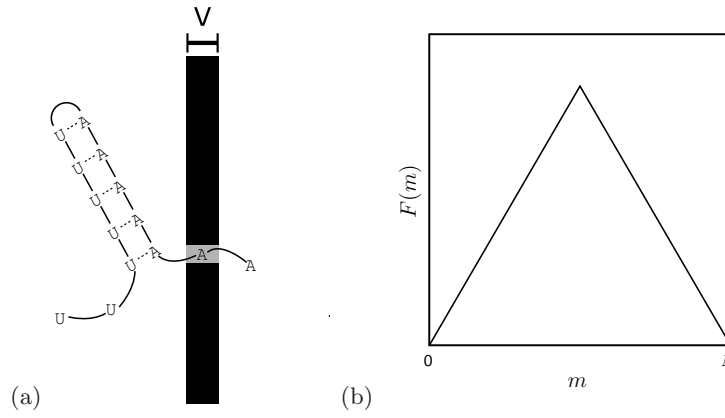
FIG. 2: Schematic of the translocation of an RNA hairpin. The RNA molecule is composed of $N/2$ A's followed by $N/2$ U's, causing all bases to be paired in the initial fold, when the molecule is completely on the cis-side of a membrane (black) with an embedded nanopore (grey). (a) The RNA is shown with one end threaded through the nanopore to the trans-side, after two bases have unbound in order to translocate. (b) Schematic of the Free Energy Landscape for translocation (Eq. 2) at the critical voltage such that $F(0) = F(N)$.

associated with RNA base stacking is small enough to prevent the opening of base pairs at typical experimental temperatures [39], we can ignore fluctuations about this ground state hairpin configuration throughout. The problem is then equivalent to unzipping a small double stranded DNA segment by translocating it through a nanopore under applied voltage [17], or more simply, unzipping double stranded DNA under a constant force [40, 41].

As successive base pairs are unzipped, the translocation process steps through a set of well-defined intermediate states labeled by the number of bases that have translocated to the trans side, m, each with a well-defined free energy, $F(m)$, where $0 \leq m \leq N$. Consider the change in free energy when the first base is translocated in Figure 2(a), $F(1) - F(0)$, where we have set the zero of free energy such that $F(0) = 0$. In order for this base to translocate, the system must overcome an energy barrier of ϵ to break the base pair, but gains an energy η from the applied voltage giving $F(1) - F(0) = \epsilon + \eta$. In fact, this is true for $F(m) - F(m-1)$ for $m \leq N/2$ since all of these bases have complementary partners on the cis-side. For $m > N/2$, we have $F(m) - F(m-1) = \eta$, since no base pairs have to be broken. The free energy for state m is thus equal to the sum of all the free energy changes to get there

$$F(m) = \sum_{i=1}^{m} \left[F(i) - F(i-1) \right], \tag{1}$$

or

$$F(m) = \begin{cases} m\epsilon + m\eta, & m \leq N/2 \\ (N/2)\epsilon + m\eta, & m > N/2 \end{cases}, \tag{2}$$

where the reduced voltage $\eta = \eta(V) = e_{eff}V$, with e_{eff} the effective charge per nucleotide and V, the voltage drop across the membrane, which is essentially equal to the applied voltage in experiments. A schematic of Eq. 2 is shown in Figure 2(b) at a voltage, η_c, such that $F(N) = F(0)$. We can see that the necessity to unzip base pairs before translocating gives a significant barrier hindering the process, unless a sufficient voltage is applied. This *one-dimensional* free energy landscape picture leads naturally to a description of the translocation process in terms of a walker moving on the landscape governed by these local energy barriers [41].

This description is however over simplified and even on this level one can see that the RNA translocation problem is more subtle and complicated than the analogy to the DNA unzipping problem in the hairpin example. In the description above we assumed there are no degenerate folds with the same energy on the cis-side. However, consider the state $m = 1$, where one base has been translated. There are now $N/2$ U's and $N/2 - 1$ A's on the cis-side, with one U left unpaired. There are $N/2$ choices for this unpaired U, thus making this state highly degenerate [42]. Despite the fact that in the initial $N/2$ translocation steps a base pair always needs to be opened for translocation to occur, the degeneracy changes with m.

For random RNA sequences it is well known that the energetic states of RNA folds are highly degenerate [43, 44] (Appendix A) and the dynamics can be complicated in a much more severe manner than described in the case of the hairpin. Figure 3(a) depicts a schematic of rna translocation for the sample sequence UAUAA in which no refolding is allowed on the trans-side. Initially this sequence can assume one of five degenerate folds, each with two base pairs on
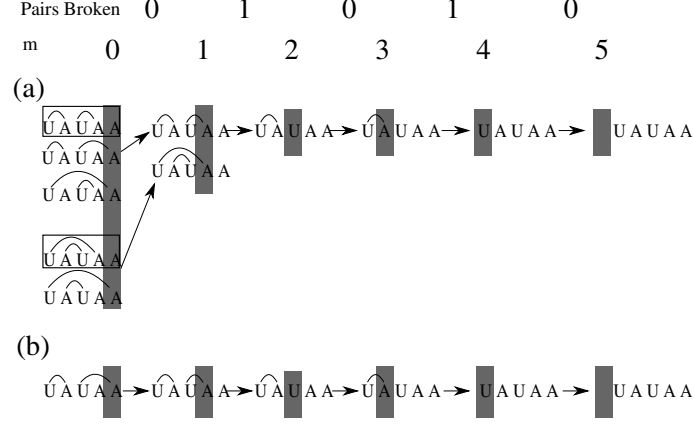
FIG. 3: Schematic of RNA translocation with no trans-side refolding. The example sequence UAUAA is depicted translocating across a nanopore (grey), which only allows the passage of single nucleotides. Secondary structure bonds between nucleotides are depicted as arcs between letters, and the net bonds broken at each step are displayed as numbers in between translocation steps. The number of bases translocated, m, is indicated above each pore. (a) The translocation process is illustrated showing all degenerate folds. Initially the complete sequence is folded on the cis-side in one of five possible folds, each with two base pair (see Figure 11). When the first nucleotide crosses the pore (first column), three of these structures require a bond breakage, while the other two do not (boxed). These three structures can revert to one of the other two (arrows) by refolding on the cis-side giving a net zero bonds broken in this step of translocation. In the second step of this process, the second nucleotide enters the pore, forcing a bond to be broken in both possible structures. Since the energetics of translocation for all possible structures reduce down to the energetics of the translocation of the topmost structure, in this study we only examine its translocation, giving a well-defined, one-dimensional energy landscape. Note that this corresponds to the quickest possible translocation, since the minimum number of base pairs have to be broken to translocate the sequence. (There are no breakages followed by refoldings). The sum of the number of bonds broken equals two, corresponding to the maximal number of base pairs possible for this structure. (b) The same process depicted in (a) for the chosen sequence.

the cis-side [45]. After the first base has been completely translocated there are now only two degenerate states. Some of the initial states require *bond rearrangement* (breakage followed by re-pairing) to collapse onto these two states, while the others do not. This makes the description of the process in terms of a one-dimensional energy landscape ill-defined within this energetic model. In fact, the energy landscape that embodies this translocation process is highly branched. At every translocation step, there are branches for each of the possible folds, with appropriate energy barriers.

Describing these energy landscapes is a formidable task and instead of considering the full multidimensional landscape, we project these degenerate states onto an effective one-dimensional landscape. To this end, we describe the translocation process by only considering the *net change* in base pairs between translocation states labeled by the one-dimensional coordinate m, representing the number of bases translocated to the trans side. By ignoring all bond rearrangements during translocation, we are effectively removing many possible intermediate energy barriers that are present in the full, highly-branched landscape. That is, within the full landscape, we always choose a trajectory with the minimum number of base pairs rearrangements (openings and closings). Note that this makes both forward and backward steps of the translocation process happen between well-defined states.

With this choice then, the RNA translocation process can be modeled as the motion of a random walker on a one-dimensional effective free energy landscape, just as in the case of modeling the dynamics of dsDNA unzipping [41]. This landscape represents the minimum number of base rearrangements that have to be undergone, and thus the dynamics of the random walkers on the landscape are only an *upper bound* to the true dynamics of the system. In what follows below, we systematically investigate the free energy landscapes and dynamics of both the no-refolding and refolding cases for random sequences drawn from the 2 letter model.

B. Calculation of Translocation Landscapes

In order to calculate the free energy landscapes of translocation, we need to find the total number of base pairs on both the cis and trans sides at each step of translocation. Since we are assuming the system reaches the ground state

at each step, we need to find the optimal fold of each of the subsequences confined to the cis and trans sides for a given m. For simplicity, we choose the simplest maximal-matching folding algorithm in order to find these optimal folds. Most notably, we ignore knotted or pseudo-knotted configurations [46].

The Nussinov folding algorithm [46], outlined in Appendix A, inputs a base sequence representing an arbitrary RNA molecule, and returns a matrix M, whose elements $M(i, j)$ represent the maximal number of base pairs allowed within the subsequence starting at base i and ending at base j. We consider the sequence to be translocating in the direction such that base N is the first to enter the pore. We can thus use this matrix M to calculate the change in the number of base pairs at each step of translocation, and by doing so ignore any internal rearrangements on either side that keep the number of base pairs fixed.

In the case of no-refolding allowed on the trans-side, $M(1, N-m)$ represents the number of base pairs on the cis-side after m bases have translocated. The energetic cost of translocating these m bases is thus the energy required to break a base pair, ϵ, times number of base pairs broken, $M(1, N) - M(1, N - m)$. Each base pair that is translocated also contributes a factor of η to the free energy, representing the voltage bias applied across the nanopore. Thus the free energy of the state where m bases have been translocated, $F(m)$ is given by

$$F_{\mathrm{no-refolding}}(m) = \epsilon(M(1, N) - M(1, N - m)) + \eta m. \tag{3}$$

If we define a variable

$$\sigma(i) = M(1, i) - M(1, i - 1) + \eta/\epsilon, \tag{4}$$

this can be rewritten as

$$F_{\mathrm{no-refolding}}(m) = \epsilon \sum_{i=N-m+1}^{N} \sigma(i) \tag{5}$$

The latter form is convenient for considering dynamics on $F(m)$ because of its resemblance to a cumulative sum of random numbers, which is a well-characterized problem (see Section IV) [34]. Figure 4(a) depicts the schematic translocation landscape and intermediate states for the sample sequence UAUAA for no-refolding on the trans-side.
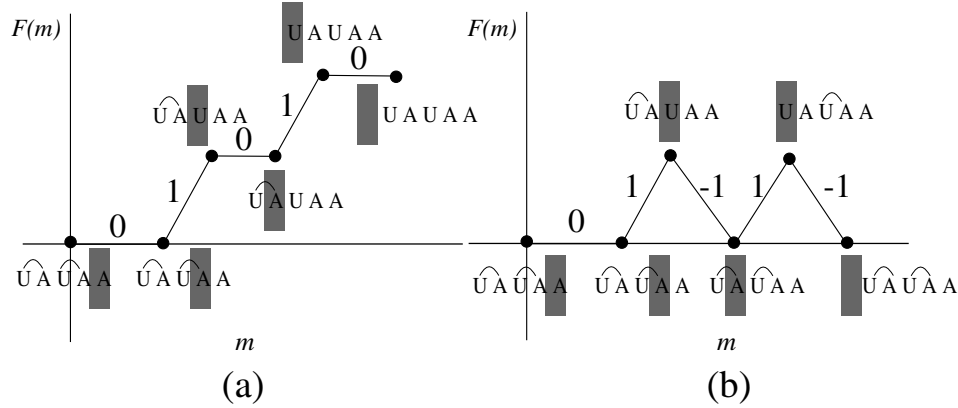

FIG. 4: Schematic free energy landscapes for translocation, $F(m)$ for (a) no-refolding and (b) refolding for the sample sequence UAUAA. Free energy is in units of ϵ. Each intermediate state of translocation is labeled with a schematic of the RNA fold (see Figure 3). The landscapes have $\eta = 0$. Between each state is labeled the net base pairs broken: (a) $\sigma(i)$, (b) $(\sigma(i) - \rho(i))$.

In the case of refolding, we must consider how many base pairs can be formed within the subsequence from base m to base N, $M(m, N)$. The difference in the number of base pairs formed between translocation steps is embodied

in the variable $\rho(i) = M(i, N) - M(i+1, N)$. For simplicity, we consider a pore too thin to accommodate any bases. Since new base pairs on the trans-side contribute a factor of $-\epsilon$ to the free energy, we have

$$F_{\text{refolding}}(m) \ = \epsilon \sum_{i=N-m+1}^{N} \sigma(i) - \epsilon M(N - m + 1, N) \tag{6}$$

$$= \epsilon \sum_{i=N-m+1}^{N} (\sigma(i) - \rho(i)), \tag{7}$$

where it is understood that $M(N+1, N) = 0$. Figure 4(b) depicts the translocation landscape and intermediate states for the sample sequence UAUAA for refolding on the trans-side.

Thus for both no-refolding and refolding, the translocation landscapes have the form of a cumulative sum. The dynamics of RNA translocation can be studied in the same way as the dynamics of DNA unzipping by studying the motion of random-walkers on these one-dimensional free energy landscapes [41].

When the cumulative sum is over uncorrelated random variables with a finite variance, the resulting energy landscape is known as a random forcing landscape since the derivative of the free energy is random [34]. As studies on random forcing energy landscapes show, the dynamics of RNA translocation as governed by FELs from the above equations are anomalous in certain regimes depending on the tilt of the landscape, η, and the base pairing energy, ϵ. In general the dynamics become anomalous where the overall tilt of the energy landscape is close to being flat. We stress that this is true, provided that the random variables $\sigma(i)$ and $(\sigma(i) - \rho(i))$ are translationally invariant, and independent.

Below we consider the dynamics on the energy landscapes of RNA translocation that we generate for randomly drawn RNA sequences from the Nussinov folding algorithm. In particular we consider the number of translocated base pairs, $m(t)$, or equivalently the displacement on the energy landscape, averaged over both disorder and thermal realizations. Numerically we perform a thermal average, denoted by $\langle \cdots \rangle$, of the dynamics over single landscapes by repeating the dynamics many times for the same landscape. A disorder average, denoted by an overline, $\overline{\cdots}$, is obtained from these thermally-averaged dynamics by averaging over many landscapes. We show that in effect the dynamics on these energy landscapes, despite the apparent complexity of the problem, are the same as dynamics on random forcing energy landscapes and exhibit the different regimes of anomalous dynamics, governed by ϵ and η, and characterized by the following properties:

- For a specific value of the voltage, where the energy landscape is on average not tilted (see Sec. III for a detailed definition) we observe Sinai diffusion where the translocation process proceeds as $\overline{\langle m(t) \rangle} \sim \log^2(t)$

- Near the above value of the voltage, the velocity of the transloction process becomes ill-defined and the translocation process proceeds as $\overline{\langle m(t) \rangle} \sim t^{\mu}$, $\mu < 1$. The exponent μ changes continuously with increasing voltage between $0 < \mu < 1$.

- Finally, outside this regime, for large enough voltages, the velocity becomes well defined with $\mu = 1$, and $\overline{\langle m(t) \rangle} \sim t$.

As discussed above, these regimes are an upper bound to the dynamics on the fully branched energy landscapes of RNA translocation.

III. RESULTS

In this section, we discuss both the free energy landscapes and dynamics of translocation for the simplified two-letter RNA model system discussed above.

For a given translocation landscape, dynamics were generated using a standard Monte-Carlo procedure [41]. For each realization, a walker was placed at position $m = 0$ of the landscape. At each time step a move $m \rightarrow m \pm 1$ was attempted with transition rates

$$w_{m \rightarrow m \pm 1} = \min\{1, e^{-(F(m \pm 1) - F(m))/k_B T}\}. \tag{8}$$

with absorbing boundary conditions at the end of the landscape ($m = N$) and reflecting boundary conditions at the beginning ($m = 0$). We choose $k_B T = 1$ so that that ϵ and η are dimensionless and take $\epsilon = 1$. The position was then recorded every $S/1000$ steps with S the total number of Monte-Carlo attempts. A thermal average of the position, $\langle m(t) \rangle$, for each realization of the translocation landscape was obtained by repeating the process ten times for each realization, and averaging the position at each of these 1000 time points.

To obtain an average over disorder the process was repeated for 1500 realizations of random two-letter RNA sequences of length $N = 4000$. The disorder averaged dynamics trajectory, $\overline{\langle m(t) \rangle}$, was calculated by averaging $\langle m(t) \rangle$ over realizations of the translocation landscapes. Below we present both the dynamics and the landscapes for a variety of tilts, η, for both the no-refolding, and refolding cases. We note that changing the value of N did not effect our results.

A. No Refolding

Equation (5) describes the free energy landscape for translocation with no trans-side refolding as a cumulative sum of the variables $\sigma(i)$ (Eq. 4). If η is chosen to be sufficiently negative, then the translocation landscape for a particular sequence realization is composed of alternating uphill and downhill segments of varying length (Figure 4(a)). In our simple model, since we are examining two-letter systems with only one type of base pair, all uphill segments have identical positive slopes, and all downhill segments have identical negative slopes.

Figure 5(a) shows a sample realization of the no-refolding translocation landscapes at the 'critical tilt' of translocation. Similar to random-forcing energy landscapes, the critical tilt is defined as the η at which the average over realizations of the landscape has no average tilt. For the 1500 realizations used in this study it is given by $\eta_c = -0.5$. Note that even though the sample landscape does not start and end at an energy of 0, the average over all realizations does. More important for understanding the dynamics are the fluctuations in the energy landscapes which are shown in Figure 5(b). These fluctuations, given by $\sqrt{\overline{(F(m) - \overline{F(m)})^2}}$, behave as $\sqrt{m}$ suggesting a structure similar to random forcing energy landscapes, where the fluctuations are known to control the dynamics [34].
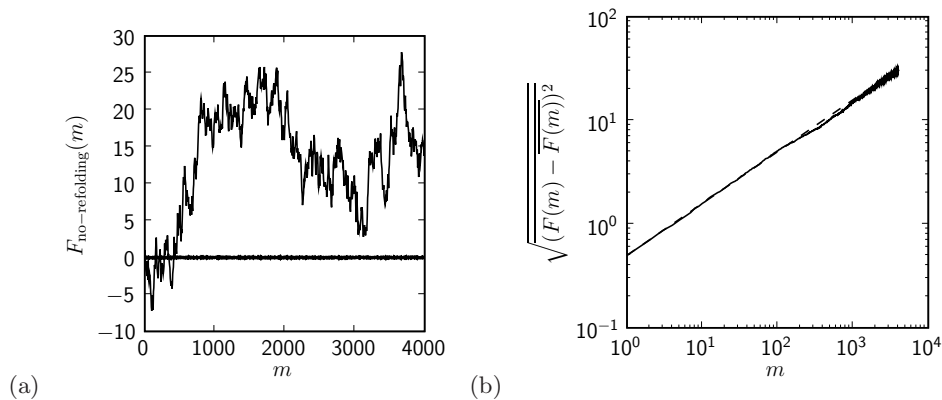

(a) (b)

FIG. 5: No-refolding translocation landscapes. A sample translocation landscape (a) is shown as a black line, and the average landscape over 1500 realizations, tilted by $\eta_c = -0.5$ is shown on the same figure (straight line). (b) The standard deviation over disorder $\sqrt{\overline{(F(m) - \overline{F(m)})^2}}$, adjusted by a tilt of -0.006 is shown on a log-log scale with $\sqrt{m}$ shown as a dashed line.

In Figure 6 we show the associated dynamics on these landscapes. Different tilts represent relative tilts to the critical η_c listed above. As a quick examination of the energy landscapes suggests, the dynamics are very similar to those exhibited by random walkers on random forcing energy landscapes. Indeed, the dynamics of no-trans-side refolding translocation at the critical tilt follows $\overline{\langle m(t) \rangle} \sim \log^2(t)$, a characteristic of Sinai dynamics. As the tilt is gradually increased, the dynamics obey a sub-linear powerlaw, $\overline{\langle m(t) \rangle} \sim t^\mu$, where $\mu < 1$, which changes continuously to $\mu = 1$ for high enough tilts.

B. Refolding

Equation (6) describes the free energy landscape for translocation allowing trans-side refolding as a cumulative sum of the variables $\sigma(i)$ (Eq. 4) and $\rho(i)$. If we set $\eta = 0$, then $F(0) = F(N) = 0$ since there will be the same number of base pairs formed in the optimal fold on the cis-side before translocation and the trans-side after complete translocation. Therefore $\eta_c = 0$ represents the 'critical tilt' for trans-side folding translocation (Figure 4(b)). We expect that at this tilt the energy landscape will be peaked on average at $m = N/2$.

Figure 7(a) shows a sample realization of the refolding translocation landscapes at the 'critical tilt' of translocation. We note that, despite the simplified model, the average landscape is similar to the translocation landscape for a random *four*-letter sequence calculated with a more sophisticated method that fully folds the RNA at each step using experimental free energy parameters ([16] Figure 3(b) inset). As before, our interest is not just the mean energy landscape but also the fluctuations about the mean. In the refolding case these must vanish at $m = N$ and the expected structure might be more elaborate than the no-refolding case. In general the fluctuations in the energy landscape will also not be symmetric about the mean value. In particular, negative fluctuations are less relevant since
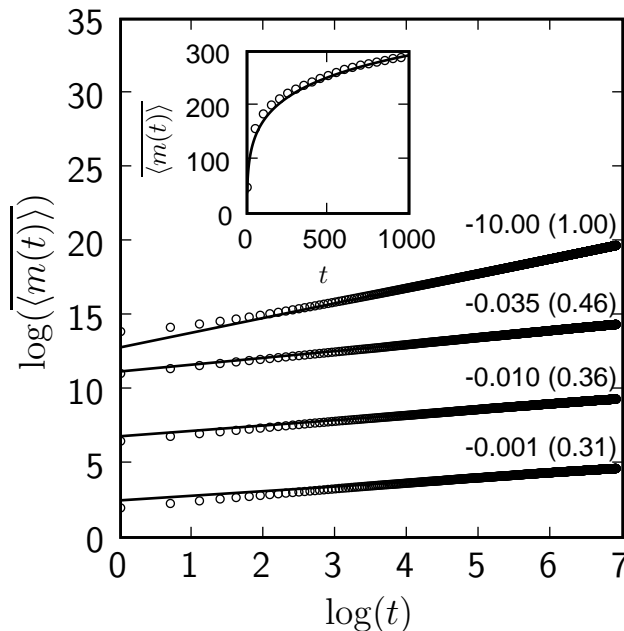
FIG. 6: Trans-side no-refolding, disorder-averaged translocation dynamics. $\log(\overline{\langle m(t)\rangle})$ vs. $\log(t)$ is plotted for four different landscape tilts (open circles, shifted for clarity). The average tilt of the landscape is shown next to each power law fit (solid line), with the exponent of the fit indicated in parentheses. The total number of steps taken, S, was $1\text{x}10^5$ for the three intermediate tilts, and $1\text{x}10^3$ for the high tilt condition. Inset: $\overline{\langle m(t)\rangle}$ vs. t for the zero-tilt landscape (open circles) with a fit to $x(t) = A + B\log(Ct)^2$ with $A = 29.6$, $B = 3.55$, $C = 5.5$, and $S = 1\text{x}10^7$. Only one out of every 50 points is shown for clarity.

they are bounded by the minimal value of the energy landscape which is zero. Despite these complications, in Fig. 7(b) we see that for a range of m, the positive fluctuations grow as $\sqrt{m}$.

The fluctuations in the energy landscape suggest that, again, the dynamics on the early part of the landscape will be similar to those of a random walker on a random forcing energy landscape. However the dynamics at large values of m might be quicker due to the constraint that $F(N) = 0$ imposed by the refolding scenario. We stress that even if this is the case, the dominating behavior will be that of the small m when anomalous dynamics are present as long as the region of anomalous dynamics grows with the system size N, which we have found to be true (data not shown).

Figure 8 shows the associated dynamics on these landscapes, where different tilts represent relative tilts to the critical value $\eta_c = 0$. We again observe the same characteristics of anomalous dynamics in the trans-side refolding translocation dynamics in the Sinai $\overline{\langle m(t)\rangle} \sim log^2(t)$ behavior in the inset. The powerlaw dynamics observed are identical to those in Figure 6, with only a slight change in exponent for a given average tilt. Since the tilts represent different relative shifts for no-refolding and refolding, this is not surprising. Once again, a large enough tilt removes the anomalous dynamics and causes $\overline{\langle m(t)\rangle} \sim t$.

In summary, for both cases of no-refolding and refolding we observe dynamics which exhibit the same features as a random walker on a random forcing energy landscape, despite the *a priori* complicated structure of the energy landscape. Namely, for small enough tilts the number of translocated bases *does not* grow as a linear function of time.

IV. DISCUSSION

Section III shows that the translocation process for both no-refolding, and refolding, exhibit anomalous dynamics for sufficiently small tilts. It is important to emphasize that due to our choice of including only base pair energetics, and our choice of degenerate folds outlined in Section II A, the simulated translocation dynamics of this model are *upper bounds* on the speed of the dynamics inherent in more sophisticated models of single-stranded polynucleotide translocation. Since anomalous dynamics are slower than $m(t) \sim t$ by definition, we expect anomalous dynamics to be observed even if additional free energy barriers, or the full degeneracy of the possible folds were included.
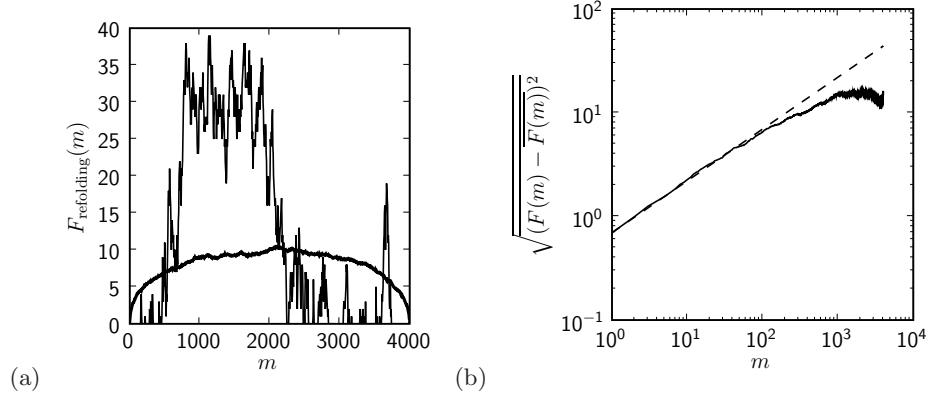
FIG. 7: Refolding translocation landscapes. A sample translocation landscape (a) is shown as a black line, and the average landscape over 1500 realizations, tilted by $\eta_c = 0$ is shown on the same figure. The sample landscape, as well as the average over all realizations, starts and ends at an energy of 0, since the maximum number of base pairs is the same when the sequence is completely on the cis-, or trans-sides.(b) The standard deviation over disorder $\sqrt{\overline{(F(m) - \overline{F(m)})^2}}$, adjusted by a tilt of -0.004 is shown on a log-log scale with $\sqrt{m}$ shown as a dashed line.
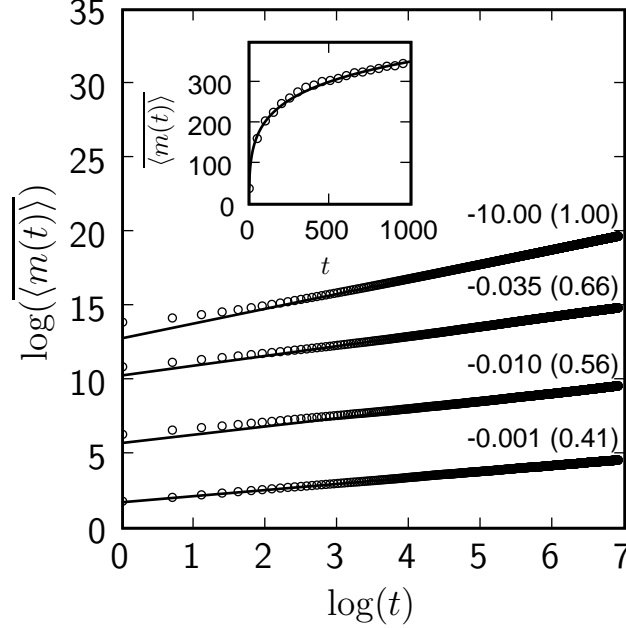

FIG. 8: Trans-side refolding, disorder-averaged translocation dynamics. $\log(\overline{\langle m(t) \rangle})$ vs. $\log(t)$ is plotted for four different landscape tilts (open circles, shifted for clarity). The average tilt of the landscape is shown next to each power law fit (solid line), with the exponent of the fit indicated in parentheses. The total number of steps taken, S, was 1×10^5 for the three intermediate tilts, and 1×10^3 for the high tilt condition. Inset: $\overline{\langle m(t) \rangle}$ vs. t for the zero-tilt landscape (open circles) with a fit to $x(t) = A + B\log(Ct)^2$ with $A = 29.6$, $B = 4.3$, $C = 5.8$, and $S = 1\times10^7$. Only one out of every 50 points is shown for clarity.

To get a better understanding of the emergence of anomalous dynamics in this model system, we rewrite Equations (5) and (6) as

$$F(m) = \sum_{i=0}^{m} \xi(i) \tag{9}$$

where $\xi(i) = \sigma(i)$ for Equation (5) and $\xi(i) = \sigma(i) - \rho(i)$ for Equation (6). The variable ξ is determined from the results of the maximal-pairing Nussinov folding algorithm (Appendix A) on a random 2-letter sequence. As stated in Sec. II A if the variables $\xi(i)$ were independently drawn from identical, say Poissonian, distributions, the free energy landscape described by Equation (9) would become a sum of random variables. Therefore, the typical fluctuations in the energy landscape would grow as $\sqrt{m}$. Such landscapes are called random forcing energy landscapes (RFEL's) since the derivative of the energy is the random variable. It is well known [34] that such energy landscapes exhibit anomalous dynamics when the average tilt of the landscape is small. Specifically, dimensional analysis implies that when the quantity $T\Delta F/\mathcal{V}$, is small, anomalous dynamics occur, where T is the temperature, ΔF is the average tilt of the energy landscape and $\mathcal{V}$ is the variance of the distribution from which ξ is drawn.

One way to understand the appearance of anomalous dynamics in random forcing energy landscapes is through a simplified model of a walker on a one dimensional lattice of traps [34]. Within this simplified picture the trap locations mimic the beginning of uphill segments, and the trap dwell times, τ, are a function of the length of the uphill segment following them. In this model, downhill segments and backward motion between traps are ignored. The trapping time in front of the uphill segment behaves as $\tau(l) \sim e^{\kappa l}$, where l is the length of the uphill segment (in dimensionless units such as the lattice spacing) and κ is roughly given by $\kappa = (\epsilon - \eta)$. The distribution of uphill lengths, l, is expected to behave as $P(l) \sim \exp(-\alpha l)$. It is straightforward to see that under these assumptions the distribution of dwell times behaves as $P(\tau) \sim \tau^{-1-\mu}$ with $\mu = \alpha/\kappa$. This broad distribution of dwell times leads naturally to the anomalous dynamics when μ is small enough. In particular, when $\mu < 1$ the average dwell time diverges leading to zero velocity. It can be shown through a straightforward analysis [34] that this leads to a displacement in time which grows as $m(t) \sim t^{\mu}$. Similar considerations lead, for example, to anomalous diffusion for $\mu < 2$. In particular, for sufficiently large values of μ, or equivalently tilts of the energy landscape, the dwell time distribution tails decay fast enough for the velocity to be well defined.

To get some understanding on the origin of the anomalous dynamics in the translocation process of our model we can look for a similar trap structure in the translocation free energy landscapes. In particular, if the distribution of uphill segment lengths across different trap sites is translationaly invariant, uncorrelated, and Poissonian, it will support the above picture which assumes that the uphill segments are drawn from identically distributed random variables.

To demonstrate that these conditions are indeed satisfied for the 1500 translocation landscapes from the folds of random 2-letter sequences of length $N = 4000$ used in this study, we computed the segment length distributions across trap sites for the no-refolding and refolding cases. For no-refolding, we identify a trap site as a start of an uphill segment. These are inevitably followed by a downhill segment. This choice is natural since uphill segments hinder translocation and 'trap' the motion, while downhill segments permit rapid translocation. Using this we record the uphill segment lengths for each trap site and study their statistics.

Figure 9(a) displays the distributions of uphill segment lengths for several trap sites for the no-refolding case. We can see that these distributions are essentially the same across these trap sites, and thus the distribution of uphill lengths across trap sites can be considered translationally invariant. Moreover, the distribution can be fitted very well with a Poisson distribution. We also note that the same analysis on downhill regions of the no-refolding landscapes offer the same conclusions (data not shown).
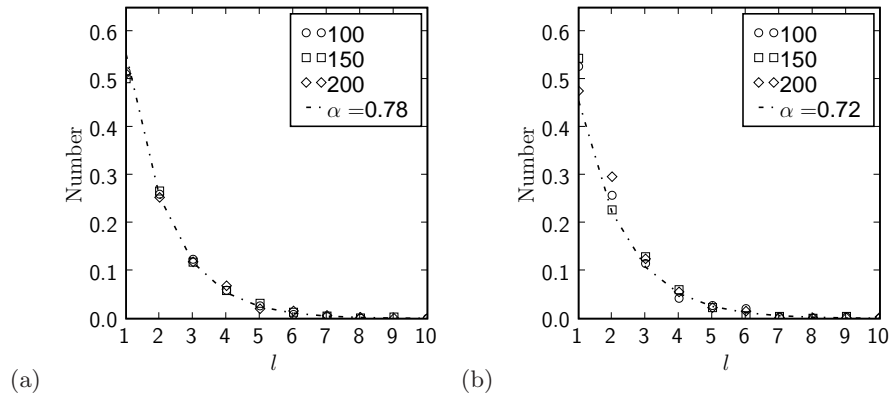

FIG. 9: Uphill segment length distributions across trap sites for (a) no-refolding, and (b) refolding. For each of the two scenarios, translocation landscapes were partitioned into trap sites by finding the locations of the beginning of uphill segments as discussed in the text. Displayed are normalized histogram heights representing the uphill length distribution for sites numbered 100, 150 and 200 (legend). These heights were fitted to a Poisson distribution $Ae^{-\alpha l}$, where l is the trap size.

In the case of refolding the definition of a trap sites is slightly modified, and consists of an uphill-flat-downhill segment triplet. The flat segments lie between an uphill and downhill segment and are due to a compensation event caused by refolding on the trans-side (Section II B). If a downhill segment immediately followed an uphill segment, a length of 0 was recorded for that flat segment. Just as in the case of no-refolding, Figure 9(b) indicates that the uphill segment length distribution for refolding landscapes is translationally invariant and Poissonian. The same is true also for the downhill segment length distributions (data not shown).

The presence of flat segments in the translocation landscapes is the primary feature distinguishing the refolding and no-refolding cases. As discussed in Appendix B, it is interesting to note that the flat segment lengths become very large, with some translocation landscape realizations composed mainly of one very large flat segment. When a walker encounters a flat segment, it undergoes diffusive motion, which is much faster than the essentially solitary motion of a walker at the base of an uphill barrier. Moreover, with a slight overall tilt, flat segments become downhill regions, while uphill segments remain uphill. Thus refolding landscapes have more effective downhill regions than no-refolding landscapes. Aiming only to get an understanding of the origin of the anomalous dynamics, we do not consider flat segments to contribute to the dwell time at a trap site, and thus leave the discussion of their distribution over trap sites to Appendix B.

To argue that the simple scenario advocated above is plausible we also have to show that there are no long-range correlations between the length of uphill segments at different sites. To do this we compute their correlation across different trap sites, separated by k sites, directly using the equation

$$C(k) = \frac{1}{N_T - k} \sum_{i=1}^{N_T - k} (\overline{(u(i)u(i+k))} - \overline{(u(i))(u(i+k))}), \tag{10}$$

where $u(i)$ represents the uphill segment length at trap site i, N_T is the total number of trap sites, and the over-line indicates a disorder average over realizations of the landscapes. Figure 10 shows that there are no significant correlations in the uphill segment lengths across trap sites for both no-refolding and refolding scenarios, with similar conclusions for the correlation among downhill segments (data not shown).
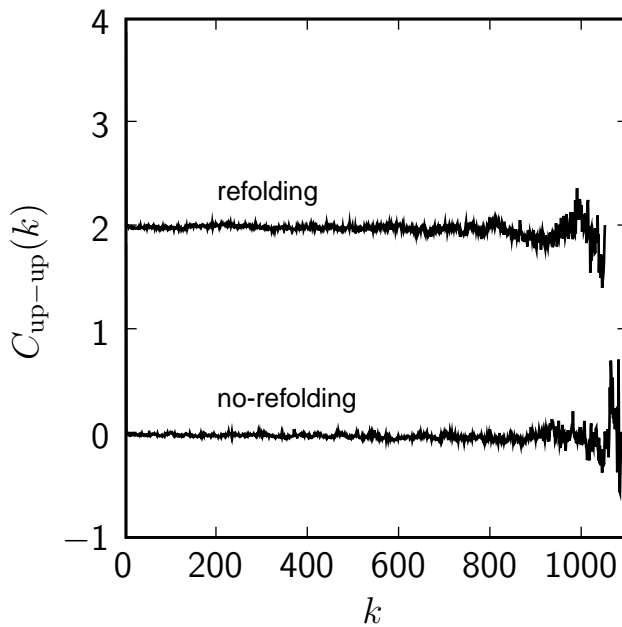

FIG. 10: Correlation function (Equation 10) for uphill segment lengths for no-refolding and refolding, shifted by 2 for clarity. In both cases there is a markedly pronounced lack of correlations.

Thus the numerics verify that the uphill segment lengths are independently drawn from identically distributed Poisson distributions making the picture presented above for the emergence of anomalous dynamics plausible. The anomalous dynamics observed in Section III are thus due to the random forcing nature of the translocation process. In particular, the no-refolding free energy landscapes seem identical to random forcing energy landscapes, despite the complicated fold of the RNA.

We have demonstrated that the dynamics associated with the translocation of RNA sequences in our model are anomalous, but we have made many simplifications. As pointed out in Section II A, we are ignoring any polymer configurational entropy barriers that might be present such as loop entropies. In addition, we have not included the excluded volume effects of realistic polymer chains. Since the anomalous dynamics we consider here are translocation-rate limiting, we expect that they would not change if these details were included. Furthermore, Figures 5(b) and 7(b) both show that the free energy fluctuations due to the base pairing effects alone, grow as $\sqrt{m}$, which dominates any logarithmic terms introduced by these other effects as mentioned above. We have also ignored any interactions the nucleotides can have with the pore itself [25]. These interactions would modify the $\sigma(i)$ variable in Equation (4) through a now sequence-dependent $\eta(i)$. It is straightforward to see that these *bounded* contributions to the energy landscape will not effect our results.

Finally we have only considered 2-letter systems, but more realistically, we can consider 4-letter systems that are closer to actual RNAs. The possible folds are more complex, but we do not expect that this would remove the anomalous behavior of the translocation dynamics. In fact, due to the similarity between the average refolding translocation landscape in Figure 7(a), and the translocation landscape for a random 4-letter sequence studied elsewhere ([16] Figure 3(b) inset), we expect our results to hold also in those systems.

V.   CONCLUSION

We have presented a simple model for the translocation of RNA through a nanopore which only allows passage of single-stranded nucleotides. At the heart of this model is the translocation free energy landscape which is calculated from the fold of a particular RNA sequence and the voltage bias applied across the nanopore.

The translocation dynamics, modeled as the motion of a random walker on these calculated translocation landscapes, was shown to display anomalous characteristics such as Sinai and sub-linear power-law regimes for sufficiently small voltage biases across the nanopore. This is despite the complicated nature of the RNA folds. In fact, this might suggest that even more sophisticated models which consider the full RNA folding dynamics should display the same anomalous characteristics.

VI.   ACKNOWLEDMENTS

APPENDIX A: NUSSINOV FOLDING ALGORITHM

In this section we discuss a modified version of the maximal folding algorithm of Nussinov and Jacobsen [46] for folding two-letter random sequences. We strongly follow the discussion in [46], but add focus to the topic of degeneracies in the possible secondary structures since it is of importance to the study of RNA translocation.

The Nussinov maximal matching algorithm returns the maximum number of base pairs possible in the secondary structure of a given one-dimensional sequence of complimentary pairing bases. In this work we focus on two-letter sequences, for example, only containing the bases A and U, where A can pair with U, but not with itself and vice versa. At the heart of the algorithm is the M-matrix, where $M(i,j)$ is equal to the maximum number of base pairs possible for the subsequence from index i to j inclusive. If N is the length of the sequence being considered, then $M(1,N)$ is the maximum number of base pairs possible in the secondary structure of the whole sequence.

The algorithm starts off with all diagonal entries of M set to 0, since bases cannot pair on themselves. It next considers all possible base pairs between adjacent bases (subsequences of length 2) to fill in the first off diagonal of M, setting elements to 1 if the bases can pair on each other, or to 0 if they cannot. The algorithm continues to sweep over all possible contiguous subsequences of increasing length, always using information about maximal pairing in the previous smaller subsequences to determine the maximal pairing in the current subsequence, and filling the entries in M.

A code listing [47] with the details of the algorithm is below. The algorithm initializes the M-matrix to be an NxN zero matrix (A), where we are using 1-based subscripting. The outer iteration is over all possible subsequence

lengths `2..N`, where `1..3` represents `1,2,3` in succession (B). Consider the subsequence from index `start` to index `end`. For this subsequence, the algorithm iterates over all indices (`intermediate`), and tests whether or not those bases can pair with base `end` (C). A vector (`pair_counts`) stores the number of base pairs that are formed when base `intermediate` pairs with base `end`, which makes use of M-matrix elements for the maximal number of bases possible in the subsequences from `start` to `intermediate-1` and `intermediate+1` to `end-1`, since these subsequences are not interrupted by the bond between bases `intermediate` and `end` (D). Once all possible pairings with base `end` are considered, and the no-pairing condition is taken into account, `M(start,end)` is set to the maximum of these stored results (E).

```
N = length(sequence)
M = zeros(N,N) #A
for sub_length in [2..N]: #B
    for start in [1..N-sub_length+1]:
        end = start+sub_length-1
        sub_sequence = sequence[start:end]
        N_sub_sequence = len(sub_sequence)
        pair_counts = zeros(N_sub_sequence,1)

        for intermediate = [1..N_sub_sequence-1]: #C
            if sub_sequence[intermediate] == sub_sequence[end]:
                pair_counts = M[start,start+intermediate-1] + 1 + \
                              M[intermediate+1,end-1] #D
        pair_counts[N_sub_sequence] = M[start,end-1]
        M[start,end] = max(pair_counts) #E
return M
```

Figure (11) depicts a schematic of this algorithm for the sample sequence UAUAA. Once the diagonal entries of M are set to 0, the algorithm considers the folding of all subsequences of length 2 (second row). There are a total of three base pairs possible, setting three values of M to 1. The last base pair between AA is not possible making $M(4,5) = 0$. When the subsequence length of 3 is considered, only one new base pairing is possible in addition to the ones obtained through subsequences of length two. This sets $M(3,5) = 1$. Note that even though this new element is compatible with the first UA being paired together, this is not counted in M until the full sequence of length 5 is considered (bottom row, middle fold). The algorithm continues in this manner, considering ever larger sebsequences until finally the full sequence is considered.

Note that M stores only the maximal number of base pairs possible, not which folds are actually present. Thus $M(1,5) = 2$, denoting 2 base pairs, yet this state is 5-fold degenerate as can be seen in the bottom row. The original Nussinov algorithm has a prescription for recovering the representative secondary structures for these maximal folds, but does not highlight the role of degeneracy in these folds. In this work, this degeneracy is effectively removed by the procedure by which the M-matrix is used to calculate the energy landscapes for translocation (Section II B).

APPENDIX B: FLAT SEGMENTS IN TRANS-SIDE REFOLDING TRANSLOCATION LANDSCAPES

In this section we present the distribution and correlation function for flat segments of the refolding translocation landscapes. As seen from Figure 12(a), the distribution of flat segments is roughly translationally invariant across different sites, as with the up segment distribution shown in Figure 9(b). We note however, that this distribution is much more complicated than the up segment distribution, and is much broader as reflected by the $\log(l)$ scale.

The flat-flat correlation function, shown in Figure 12(b), is negative for large length scales. A given translocation landscape has a finite length, which is equal to the length of the RNA molecule, N. We see from the distribution of flat segment lengths, that these lengths can approach 2000, which is half of the allowed N for a landscape. We thus interpret the negative correlations as being a result of these two effects: if a large flat segment occurs, there is not room in the translocation landscape for another one of the same length to occur, so small ones occur leading to negative correlations.

However, as discussed in Section IV, these properties of the flat segments do not influence our interpretation of refolding translocation landscapes as random forcing energy landscapes, and our conclusion that this results in
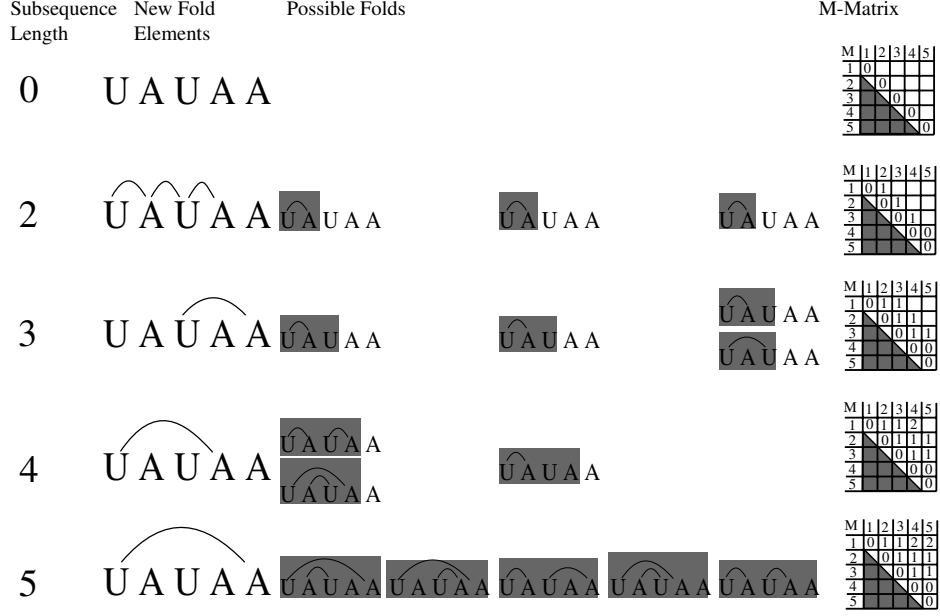
FIG. 11: Schematic of the intermediate states in the Nussinov folding algorithm applied to the example sequence UAUAA. Successive rows represent increasing values of sub_length, the subsequence length being considered. The second column indicates new secondary structures that can be formed at this subsequence length, with base pairings denoted by arcs. The next three columns represent the folds including these new elements with grey boxes indicating the subsequence that is folded. Degeneracies are indicated with stacked folds sharing the same subsequence box, except in the last row where these are shown horizontally. At each step, the M-matrix is filled in (last column), and since $M(i,j) = M(j,i)$ only the values for the top half are shown.
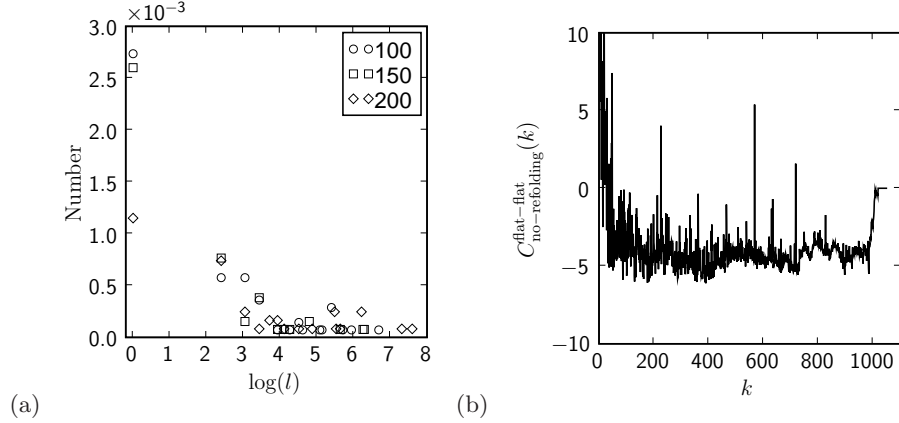

(a) (b)

FIG. 12: Refolding flat segment length distribution (a), and correlation function (b). In (a), just as in Figure 9(b), the flat segment distribution is plotted for the same sites, but versus $\log(l)$. In (b), the flat-flat correlation function (Equation 10) is plotted.

anomalous dynamical regimes.

[1] S. Takyar, R. P. Hickerson, and H. F. Noller, Cell **120**, 49 (2005).
[2] O. Namy, S. Moran, D. Stuart, R. Gilbert, and I. Brierley, Nature **441**, 244 (2006).

[3] S. Simon, C. Peskin, and G. Oster, Proc. Natl. Acad. Sci. U. S. A. **89**, 3770 (1992).

[4] M. Akeson, D. Branton, J. J. Kasianowicz, E. Brandin, and D. W. Deamer, Biophys. J. **77**, 3227 (1999).

[5] J. L. Li, M. Gershow, D. Stein, E. Brandin, and J. A. Golovchenko, Nat. Mater. **2**, 611 (2003).

[6] J. J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer, Proc. Natl. Acad. Sci. U. S. A. **93**, 13770 (1996).

[7] D. W. Deamer and D. Branton, Accounts Chem. Res. **35**, 817 (2002).

[8] D. Fologea, M. Gershow, B. Ledden, D. S. McNabb, J. A. Golovchenko, and J. Li, Nano Lett **5**, 1905 (2005).

[9] M. M. Inamdar, W. M. Gelbart, and R. Phillips, Biophys. J. **91**, 411 (2006).

[10] K. Jeang and V. Yedavalli, Nucleic Acids Res. **34**, 4198 (2006).

[11] W. Wintermeyer, F. Peske, M. Beringer, K. B. Gromadski, A. Savelsbergh, and M. V. Rodnina, Biochem. Soc. Trans. **32**, 733 (2004).

[12] S. C. Blanchard, H. D. Kim, R. L. Gonzalez, J. D. Puglisi, and S. Chu, Proc. Natl. Acad. Sci. U. S. A. **101**, 12893 (2004).

[13] F. Vanzi, Y. Takagi, H. Shuman, B. S. Cooperman, and Y. E. Goldman, Biophys. J. **89**, 1909 (2005).

[14] A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, and D. Branton, Phys. Rev. Lett. **90**, 238101 (2003).

[15] U. Gerland, R. Bundschuh, and T. Hwa, Phys. Biol. **1**, 19 (2004).

[16] R. Bundschuh and U. Gerland, Phys. Rev. Lett. **95**, 208104 (2005).

[17] S. Kotsev and A. Kolomeisky, J. Chem. Phys. **125**, 084906 (2006).

[18] Y. Kafri, D. K. Lubensky, and D. R. Nelson, Biophys. J. **86**, 3373 (2003).

[19] S. M. Bezrukov, I. Vodyanoy, and V. A. Parsegian, Nature **370**, 279 (1994).

[20] A. Meller, L. Nivon, E. Brandin, J. Golovchenko, and D. Branton, Proc. Natl. Acad. Sci. U. S. A. **97**, 1079 (2000).

[21] A. Meller, L. Nivon, and D. Branton, Phys. Rev. Lett. **86**, 3435 (2001).

[22] A. Meller, J. Phys.-Condes. Matter **15**, R581 (2003).

[23] J. Mathe, A. Aksimentiev, D. R. Nelson, K. Schulten, and A. Meller, Proc. Natl. Acad. Sci. U. S. A. **102**, 12377 (2005).

[24] T. Z. Butler, J. H. Gundlach, and M. A. Troll, Biophys. J. **90**, 190 (2006).

[25] D. K. Lubensky and D. R. Nelson, Biophys. J. **77**, 1824 (1999).

[26] S. Matysiak, A. Montesi, M. Pasquali, A. B. Kolomeisky, and C. Clementi, Phys. Rev. Lett. **96**, 118103 (2006).

[27] W. Sung and P. Park, Phys. Rev. Lett. **77**, 783 (1996).

[28] M. Muthukumar, J. Chem. Phys. **111**, 10371 (1999).

[29] J. Chuang, Y. Kantor, and M. Kardar, Phys. Rev. E **65**, 011802 (2002).

[30] Y. Kantor and M. Kardar, Phys. Rev. E **69**, 021806 (2004).

[31] K. Luo, I. Huopaniemi, T. Ala-Nissila, and S. Ying, J. Chem. Phys. **124**, 114704 (2006).

[32] M. Muthukumar and C. Kong, Proc. Natl. Acad. Sci. U. S. A. **103**, 5273 (2006).

[33] R. Metzler and J. Klafter, Biophys. J. **85**, 2776 (2003).

[34] J. P. Bouchaud, A. Comtet, A. Georges, and P. Ledoussal, Ann. Phys. **201**, 285 (1990).

[35] J. Mathe, H. Visram, V. Viasnoff, Y. Rabin, and A. Meller, Biophys. J. **87**, 3205 (2004).

[36] T. Ambjornsson and R. Metzler, Phys. Biol. **1**, 77 (2004).

[37] F. Krzakala, M. Mezard, and M. Muller, Europhys. Lett. **57**, 752 (2002).

[38] O. Flomenbom and J. Klafter, Phys. Rev. E **68**, 041910 (2003).

[39] R. Blossey and E. Carlon, Phys. Rev. E **68**, 061911 (2003).

[40] D. K. Lubensky and D. R. Nelson, Phys. Rev. E **65**, 031917 (2002).

[41] J. D. Weeks, J. B. Lucks, Y. Kafri, C. Danilowicz, D. R. Nelson, and M. Prentiss, Biophys. J. **88**, 2752 (2005).

[42] Contrast this to the unzipping of a DNA hairpin under constant force where the tension on the strands leaves only one state indexed by how many base pairs have been unzipped to that point.

[43] P. de Gennes, Biopolymers **6**, 715 (1968).

[44] R. Bundschuh and T. Hwa, Phys. Rev. E **65**, 031903 (2002).

[45] Note that we are assuming that each of these five states has the same free energy, even though the loops formed are of different size in each state and thus will have a different length-dependant loop entropy. See for example [48, 49]. However, as argued above, configurational entropies represent an insignificant barrier in the translocation process and will be overshadowed by base pairing effects and can thus be ignored.

[46] R. Nussinov and A. Jacobson, Proc. Natl. Acad. Sci. U. S. A. **77**, 6309 (1980).

[47] All programs for this work were written in Python, version 2.5 (http://www.python.org), using numpy (http://www.scipy.org) as a numerical routines library, and matplotlib (http://matplotlib.sourceforge.net/) for graphics plotting and fitting. When speed was necessary, parts of the python modules were written as C++ extensions using the Simplified Wrapper and Interface Generator (http://www.swig.org/). All of these packages and the python language itself are open source, and thus freely available.

[48] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).

[49] M. Fisher, J. Chem. Phys. **45**, 1469 (1966).